

Defining (Artificial) Intelligence

Correct descriptions.



S2VT: A man is doing stunts on his bike.



S2VT: A herd of zebras are walking in a field.

Relevant but incorrect descriptions.



S2VT: A small bus is running into a building.



S2VT: A man is cutting a piece of a pair of a paper.

Getting
data

Implement in
production

Data due
diligence

Make business case,
present value

Data
cleaning

Model
comparison

Model 1

Model 2

Predictive
modelling

Model
 N

Regression vs classification

- Regression: metrics like RMSE and R^2
- Classification: expected value, ROC curve

Binary classification is the simplest case, for which the classes often are simply called 'positive' and 'negative'

How to measure performance in a model: crossvalidation

Original Data



Build Model With

CV Group #1



CV Group #2



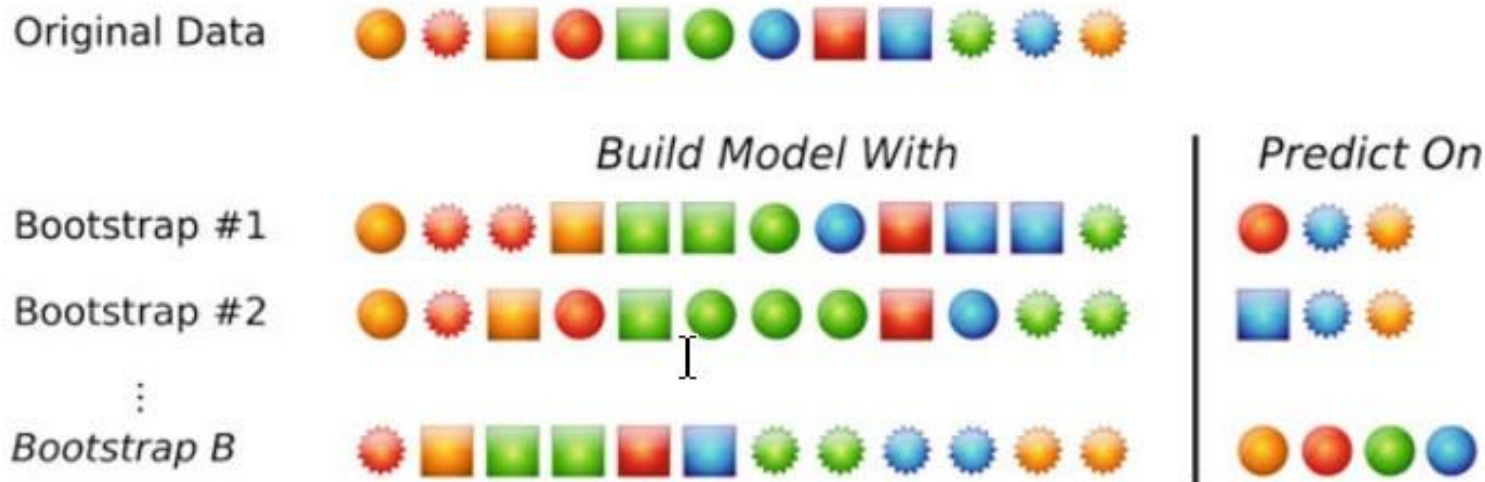
CV Group #3



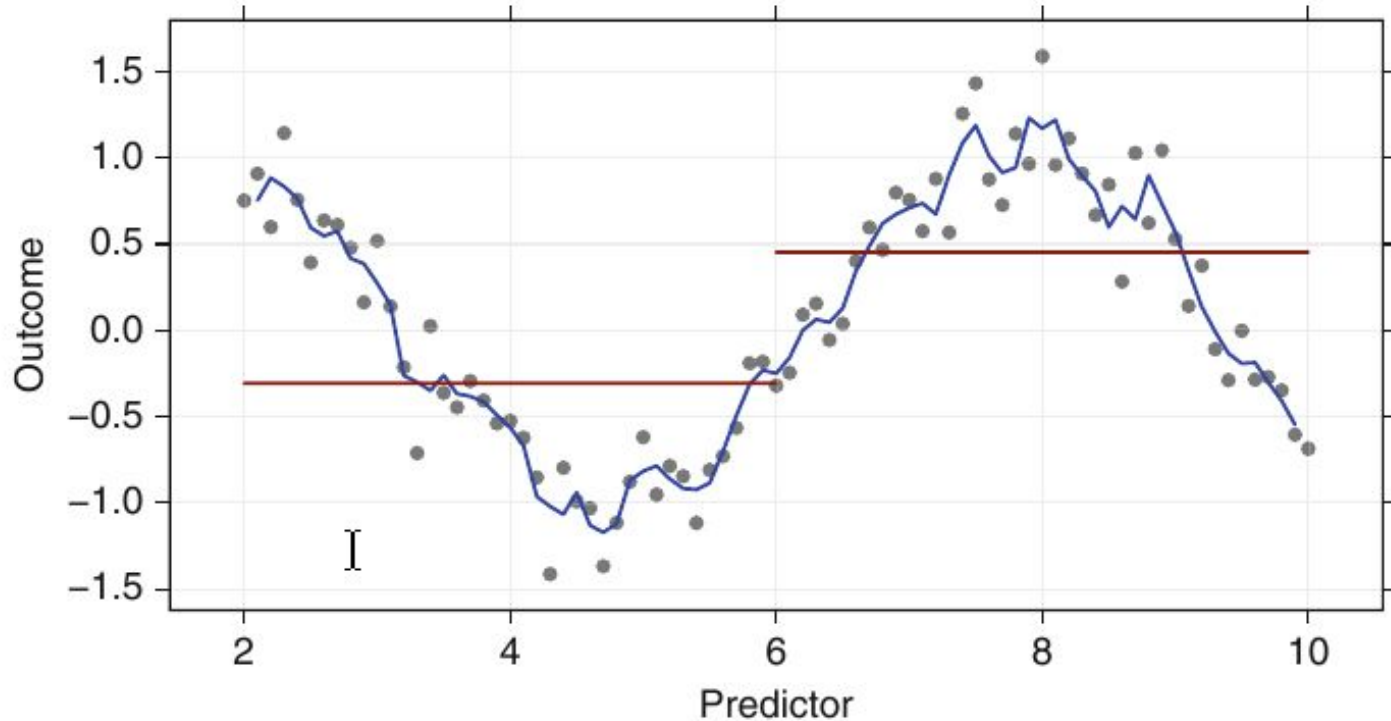
Predict On



How to measure performance in a model: Bootstrap



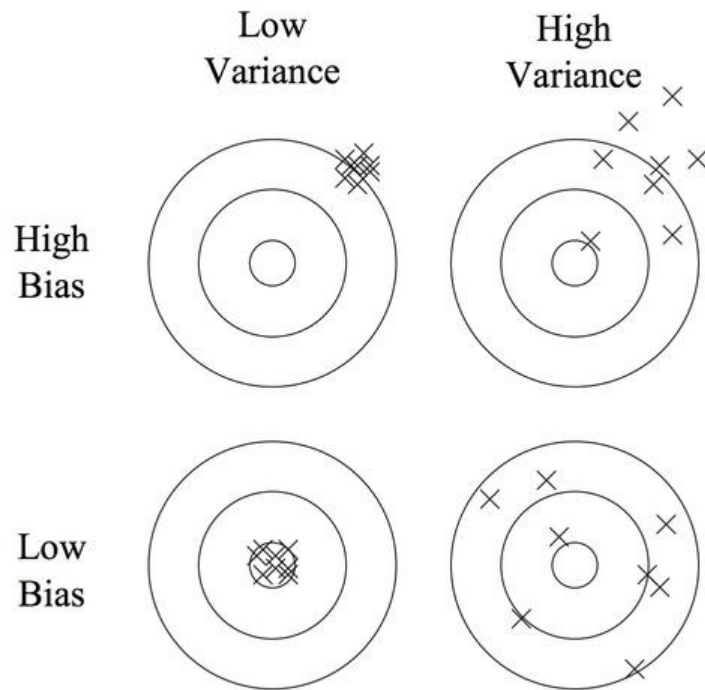
Over-fit, variance/bias dilemma



Bias-variance dilemma

- Helps to understand why simple learners outperform complex ones sometimes
- Helps to understand why model ensembles outperform single models
- Helps to understand and avoid overfitting

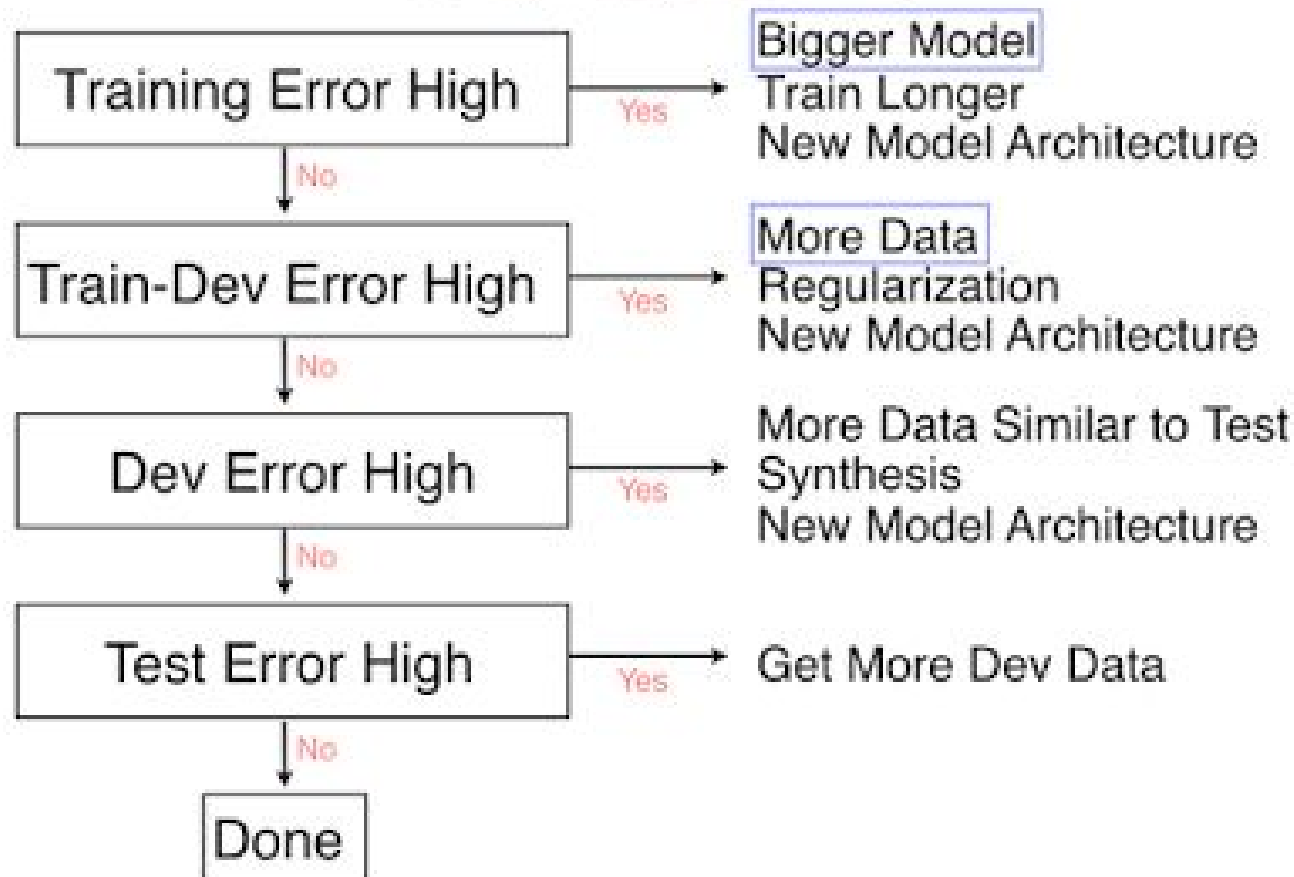
Over-fit, variance/bias dilemma



Why do we get overfitting?

Because training error is a bad estimation of generalization error

The Nuts and Bolts of Building Applications Using Deep Learning



FEATURE ENGINEERING IS THE KEY

MORE DATA BEATS A CLEVERER ALGORITHM